PREBIOTIC ASSEMBLY OF NUCLEIC ACIDS

A Thesis
Presented to
The Academic Faculty

by

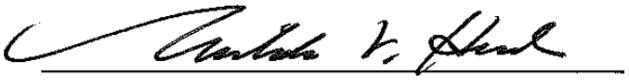
Michael C. Chen

In Partial Fulfillment
of the Requirements for the Degree
Bachelor of Science in the
School of Chemistry and Biochemistry

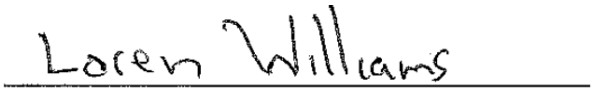
Georgia Institute of Technology
May 2012

PREBIOTIC ASSEMBLY OF NUCLEIC ACIDS

Approved by:



Dr. Nicholas V. Hud, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology



Dr. Loren Dean Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. William J. Baron
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: May 4, 2012

The ideal scientist is one who thinks like a poet, but works like a bookkeeper.

-E.O. Wilson

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. Nick Hud, who over the years helped me develop my passion for science. The opportunity to work in his lab on the chemical origins of life was an eye-opening experience that will continue to influence my career for many years to come. I would also like to thank all of the mentors, friends, and colleagues I developed in the Hud lab—Aaron Engelhart, Irena Mamajanov, Brian Cafferty, Ragan Buckley, Ford Lannan, Denise Okafor, Brandon Laughlin, and Tim Lenz—they not only helped me with my science, but also made me look forward to coming into lab every day.

I would also like to thank the programs that made it possible for me to work in the Hud Lab, including the Barry M. Goldwater Scholarship, the IBB Petit Undergraduate Research Scholarship, the Presidential Undergraduate Research/Travel Awards, the Richard Fink Research Initiation Award, the Hypercube Scholar Award, the Outstanding Undergraduate Reseearcher in the College of Sciences Award, and the Dr. William H. Eberhardt Scholarship

TABLE OF CONTENTS

LIST OF FIGURES

SUMMARY

Since the discovery of catalytic RNAs called ribozymes, there has been speculation that RNA was the first biopolymer responsible for the advent of life on earth. Life depends on the information storage provided by DNA and the biocatalysis provided by proteins. The RNA world hypothesis proposes that RNA fulfilled both of these functions in early life forms, acting both as a gene and an enzyme. However, efforts to demonstrate a prebiotically plausible synthesis of nucleic acids such as RNA have proven exceptionally challenging.

One challenge of synthesizing RNA stems from the inability to find an efficient pathway to nucleoside synthesis, the monomer of nucleic acids. Termed "the nucleoside problem," the canonical nucleobases guanine, adenine, cytosine, thymine, and uracil do not form a glycosidic linkage with ribose to a significant extent. In this thesis, work is presented showing the possibility that alternate sugars, such as ribulose, dried with adenine can form adenosine, albeit in trace amounts.

Additionally, this thesis proposes a model for base-pair selection and nucleic acid polymerization. Since short nucleic acids, or oligonucleotides, have a persistence length of 2-4 nucleotides, oligonucleotides self-cyclize before linear polymerization. Intercalation increases the persistence length of oligonucleotides to ~100 nucleotides, mediating the polymerization of nucleic acids. Intercalators also select for specific base pairs to intercalate, mediating intercalation among base pairs that topologically fit the planar, aromatic face of the intercalator.

CHAPTER 1

INTRODUCTION

1.1 THE ORIGIN OF LIFE

The central dogma of molecular biology provides a general blueprint for cellular function in all organisms: deoxyribonucleic acid (DNA) is first transcribed into ribonucleic acid (RNA) and then translated into proteins.[1] When considering the origins of life, the two cellular functions crucial to life are the information storage that DNA performs and the biocatalysis that proteins perform. DNA functions as a long-term information storage biopolymer and encodes for the proteins that catalyze metabolic pathways such as the citric acid cycle. Without protein enzymes, life would be unable to catalyze reactions such as those responsible for nucleic acid synthesis. Without DNA, life would be unable to store and pass on the genetic information that encodes for protein synthesis. Thus, an apparent paradox that is analogous to the chicken or the egg question arises: what came first, DNA or protein?[2]

In the seminal Miller-Urey spark discharge experiment, amino acids were synthesized under conditions thought to replicate the atmosphere of early-Earth.[3] Subsequently, the prebiotic synthesis of various nitrogenous bases including adenine from hydrogen cyanide as well as sugars from the formose reaction were demonstrated, making the synthesis of both proteins and nucleic acids prebiotically feasible.[4,5] Many theories have attempted to address the nucleic acid- or protein-first paradox. The metabolism-first school of thought hypothesized that life arose from prebiotic metabolic cycles catalyzed by inorganic or organic molecules, such as mineral surfaces.[6,7] Mineral

surfaces, such as pyrite proposed by Wachtershauser, would have provided catalytic sites that function as primitive enzymes similar to proteins.[6] Due to little experimental evidence however, metabolism-first theories have not held up to scrutiny.[8,9] Many metabolism-first theories propose that primitive metabolic cycles were formed by a sequential series of discrete reactions catalyzed by the same catalyst, which is problematic because the diversity of metabolites produced in these discrete reactions requires a diversity of catalytic surfaces not found outside of protein enzymes.[8] Additionally, since total percent yield of product for a sequential series of reactions is

$$\text{Total percent yield} = (\text{percent yield of step})^{\text{number of steps}},$$

each turn of a prebiotic metabolic cycle would generate fewer and fewer metabolites until all metabolites are exhausted.[10]

1.2 THE RNA-WORLD HYPOTHESIS

In the central dogma of molecular biology, RNA was relegated to the role of an intermediate polymer that relayed information from DNA to protein.[1] If nucleic acids only serve as information storage polymers, as the central dogma suggests, then a nucleic acid-first origins theory would lack the necessary catalytic capabilities for a self-sustaining biological system with emergent properties. The discovery of catalytic RNA within biological systems, particularly the ribosomal RNA which mediates peptidyl transferase activity, demonstrated not just the versatility of RNA and nucleic acids, but also supported the hypothesis that RNA was important to the emergence of life.[11] Furthermore, ribozymes have been shown to catalyze a wide array of reactions, from the cleavage of RNA to DNA transcription.[12-13] Due to its catalytic potential and inherent

information storage capability, RNA holds many of the key characteristics necessary for the development of prebiotic molecular systems capable of self-replication and evolution. Thus, developing chemical routes to the synthesis of RNA or RNA-like polymers is important to understanding how life emerged.

Assuming that RNA or a polymer similar to RNA was the first carrier of information relevant to life, a demonstration of a plausibly prebiotic *de novo* synthesis of RNA is required. However, the manner in which RNA or nucleic acids emerged in the prebiotic world remains unclear.[14] A nitrogenous base, a sugar, and a phosphate group are the constituents of a monomeric unit of a nucleic acid polymer. Discovering prebiotically feasible pathways to the synthesis of these constituents is crucial to the verification of the RNA world hypothesis. Then, these monomeric units must be polymerized to form polymers that have the ability to replicate and store information. Routes to nucleic acid synthesis are constrained by the limits of known prebiotic conditions and chemical availability. Substantial progress has been made on developing pathways to the synthesis of purines and pyrimidines.[15-16] Synthesizing a β-nucleoside and polymerizing phosphoester linked nucleosides have nevertheless been stumbling blocks for many origin of life theories such as the RNA-world hypothesis.[17] This thesis presents an investigation of the nucleosidation reaction between adenine and several types of sugars. Additionally, this thesis proposes a model for the prebiotic replication of oligonucleotides as well as nature's selection of the Watson-Crick base pair.

CHAPTER 2

PREBIOTIC SYNTHESIS OF ADENOSINE WITH ALTERNATIVE SUGARS

2.1 INTRODUCTION

By providing pathways that feasibly fit with the predicted conditions of early-Earth, the community gains insight into the emergence of functional biopolymers that gave rise to life. While the RNA world hypothesis and its offshoots paint a great picture, work showing the feasibility of the prebiotic emergence of nucleic acids does not currently exist. Before the formation of nucleic acids can even occur, the monomers of nucleic acids--nucleosides--must be present for there to be anything to polymerize. Despite this requirement, no group has satisfactorily demonstrated the prebiotic synthesis of nucleosides. The issue stems from the difficulty of linking a nitrogenous base to a sugar, the two components that make up a nucleoside.[18] This reaction produces water, which is energetically unfavorable in aqueous solutions.[19] The difficulty of N-glycosylation of nucleobases arises from the susceptibility of the glycosidic bond between ribose and a free nucleobase to hydrolytic cleavage. Additionally, the N1 electrons of many nucleobases are delocalized into conjugated π orbitals making the formation of a glycosidic bond kinetically unfavorable.[20]

Previously, Orgel and colleagues conducted dry heating reactions containing ribose and purine nucleobases in the presence of certain divalent metal ions and showed that purine nucleosides were formed such as inosine and adenosine.[21-22] The low yield of nucleoside (<1%) nevertheless cast doubt on the prebiotic relevance of these simple dry

heating reactions. Furthermore, adenine was the only canonical nucleobase shown to add onto ribose in a drying heating reaction, whereas guanosine, cytidine, and uridine were not detected. Since the electrons of N1 in guanine are not delocalized, unlike its pyrimidine counterparts, guanine should have similar reactivity compared with adenine. The solubility of guanine, however, is far less than that of adenine and is likely the reason guanine does not readily react with ribose. On the other hand, pyrimidines such as cytosine have never been shown to add onto ribose under prebiotically relevant conditions. The lack of reaction is attributed to the previously mentioned delocalization of the electrons of N1.

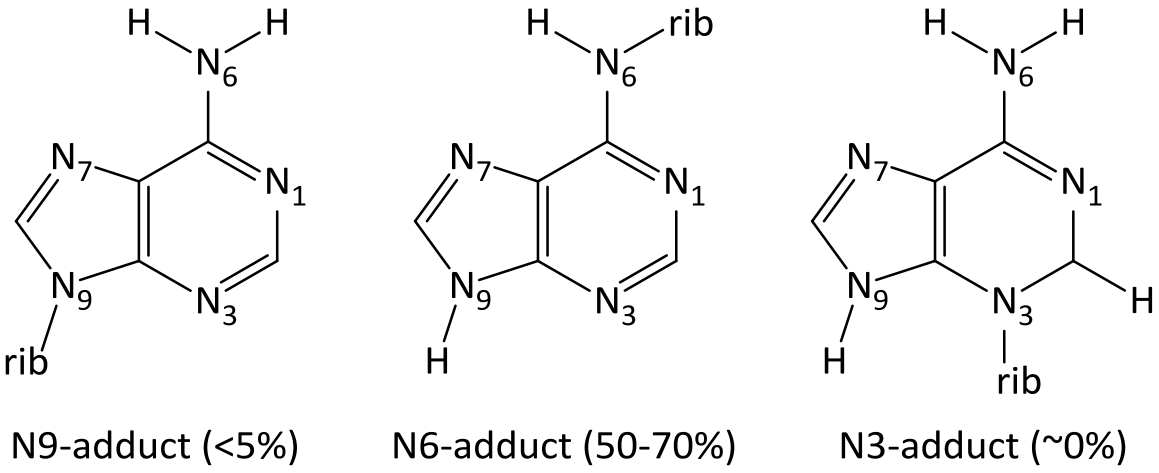


Figure 2.1. The three potential sugar-nucleobase adducts as a result of heat drying with adenine, ribose, and various divalent metal ions, as performed by Fuller et. al.[22,23] The only products were later confirmed to be the N9- and N6-adduct.[24]

Past studies from the Hud lab have shown that certain non-canonical pyrimidine nucleosides can also be synthesized through dry heating using nucleobases such as 2-pyrimidinone.[19-25] Due to the nucleophilicity of N1 and lack of electron delocalization in the structure, the reactivity of 2-pyrimidinone with ribose in dry heating reactions is significantly higher than the canonical nucleobases (~12%).[19] The difficulty of the N-

glycoslyation step has driven others to bypass the step altogether; Sutherland and colleagues suggested synthetic routes that bypassed adding the nucleobase to the sugar and focused on building the nucleobase piece-by-piece on the sugar from smaller precursors such as 2-aminooxazole.[26-27] However, Sutherland and colleagues' approach requires a specific order of addition of reactants, which would otherwise react with each other, and requires high energy intermediates of questionable prebiotic relevance.[28]

The problem with nucleosidation reactions does not just lie with the reactivity of the nucleobase, but also with the sugar.[11] The formose reaction, which involves the polymerization of formaldehyde to form a number of important organic molecules including glycolaldehyde, glyceraldehyde, and sugars, is the model prebiotic reaction that is hypothesized to have produced the supply of sugar on the early Earth.[11] The formose reaction has not been shown to produce a significant amount of ribose, which is often created in yields less than 0.1%,[29] while other aldopentoses and ketopentoses are produced in much large yields.[30]

In our view, a prebiotically feasible pathway to nucleosides has yet to be demonstrated. This chapter expands upon the investigation first conducted by Fuller et. al. of the use of alternative sugars in condensation reactions with nitrogenous bases.[22] A profile of the reactivity of sugars other than ribose to nucleobases like adenine can add support to the prebiotic relevance of nucleic acids with alternative sugars. Additionally, alternative sugars may undergo isomerizations under the heat drying cycles used to make nucleosides, which may be a way to bypass the low yield of ribose in the formose reaction. Since Decker et. al. found that ketopentoses were produced in 20% yield in a formose reaction catalyzed by calcium acetate in base, we focus specifically on studying

the reactive properties of L-ribulose.[29] Even though these alternative sugar reactions produce nucleoside at the same efficiency, the potential that multiple sugars can be used to synthesize the nucleosides found in life is highly encouraging and is a step toward demonstrating the prebiotic feasibility of prebiotic nucleoside synthesis.

2.2 EXPERIMENTAL PROCEDURES

2.2.1. Materials

Adenine, 9-β-D-ribofuranosyladenine, and D-ribose were purchased from Sigma Aldrich and used as is. L-ribulose and 9-β-L-ribulofuranosyladenine were supplied as gifts from Professor Ramanarayanan Krishnamurthy at The Scripps Research Institute. Additional L-ribulose was purchased from zuChem.

2.2.2. Heat drying reactions

Reactions were performed under three conditions. The first reaction mimics Orgel and colleagues' heat drying reaction conditions: 5 mM adenine, 75 mM D-ribose, 25 mM $MgCl_2$, and 50 mM $MgSO_4$ in 200 μL of dH_2O. The second reaction condition contained 5 mM adenine and 10 mM D-ribose adjusted to pH 2.1 by HCl. The third reaction contained 5 mM adenine and 10 mM D-ribose buffered at pH 6.3 by 50 mM sodium cacodylate. Reactions were mixed and pipetted onto 65 mm Pyrex watch glasses. The watch glasses were incubated for 2 hrs at 75 °C and 2.25 hrs at 100 °C (Orgel reaction). Reactions were then resuspended in 200 μL water after heat drying. For pH 2.1 reaction conditions, the acidity was neutralized by resuspending with 2 eq of NaOH. The samples were then filtered with a 2 μ filter and then speed-vacuumed to dryness.

2.2.3. LC-MS analysis of products

Samples from 2.2.2 were then resuspended in a solution of 80% acetonitrile/20% water to prevent strong solvent peak-splitting. Samples were chromatographed on an Agilent 1200 at an ambient temperature with an isocratic flow of 80% acetonitrile and 20% water. Formic acid (0.1% of mobile phase) was added to increase ionization efficiency for mass spectrometry. The column was a Phenomenex Luna NH_2 (150 mm × 4.60 mm × 3 μ), which required ~3 hours of equilibration before running samples and ~30-40 minutes of equilibration between each sample run. Adenine and adenine adducts were monitored by UV-VIS at 260 nm.

2.3 RESULTS AND DISCUSSION

Dry heated samples of D-ribose and adenine under the Orgel conditions were chromatographed to replicate the results reported by Fuller et. al.[22] In Figure 2.2, adenine was found to add to ribose in the N9 position at a yield of <1%. As reported by Orgel and colleagues, yields of the N6 adduct were found to be significantly higher than the N9 adduct. The three peaks labeled N6 adduct in Figure 2.2 have not been fully characterized, but their masses correspond to a condensation between adenine and ribose. Previous NMR characterization of products obtained by Fuller et. al. provided evidence that these peaks correspond to α and β isomers of 6-D-ribofuranosyladenine.[24] Bean et. al. repeated the condensation reaction between ribose and adenine under pH 2.1 conditions, and obtained similar yields of <1% of adenosine.[19]

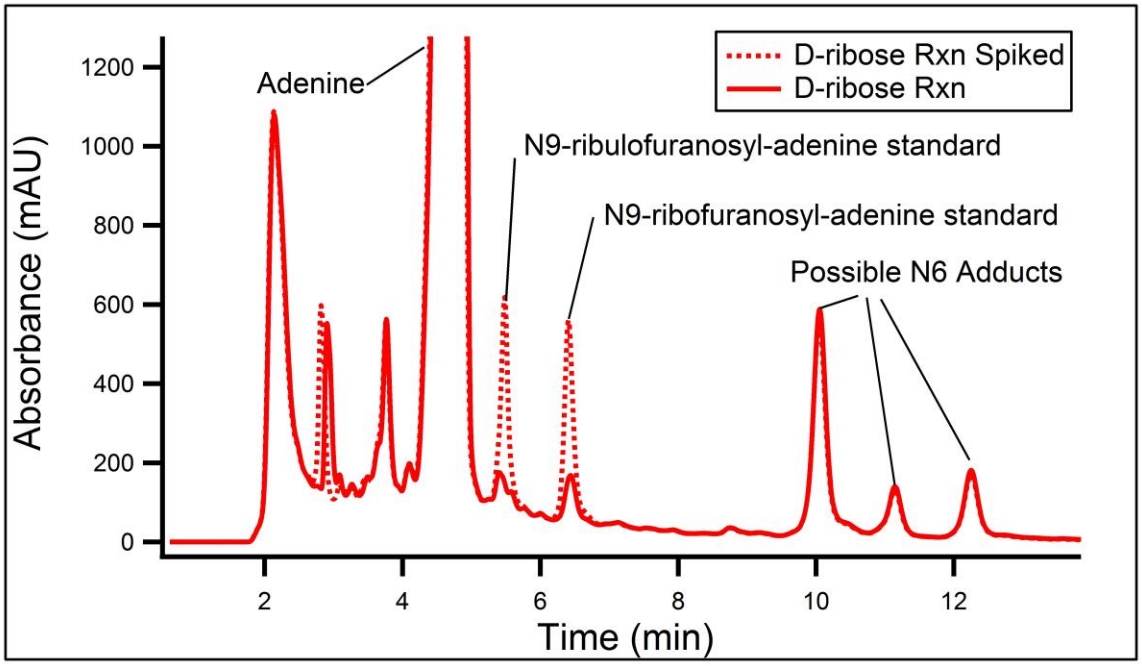


Figure 2.2. A chromatogram of an Orgel-type reaction between D-ribose and adenine. The trace spiked with 9-β-D-ribofuranosyladenine is overlaid on the trace that is not spiked. The peaks labeled adenosine match the retention time and mass of 9-β-D-ribofuranosyladenine. The peaks labeled N6 adduct corresponded to the mass of a condensation between ribose and adenine. The position of the ribose adduct was not characterized; however, previous work has shown that these peaks are likely a combination of α and β isomers of ribose N6 adducts to adenine.[24]

Since the yield of 9-β-D-ribofuranosyladenine was <1% in the Orgel-type heat drying reactions, we searched for another sugar that would theoretically be more reactive than ribose. In other words, a sugar that better stabilizes the carbocation transition state at the C1 position where nucleosidation takes places should be able to add to the N9 position of adenine in greater yield. Fuller et. al. (1972) previously investigated condensation of adenine with D-arabinose, D-xylose, and D-lyxose, discovering that yields were similar to the yields of ribose.[22] As discussed in Figure 2.3, since the furanose form of ribulose would form a tertiary carbocation, compared to the secondary carbocation formed by ribose during nucleosidation, ribulose was predicted to have a greater reactivity towards condensation compared to pentoses such as ribose.

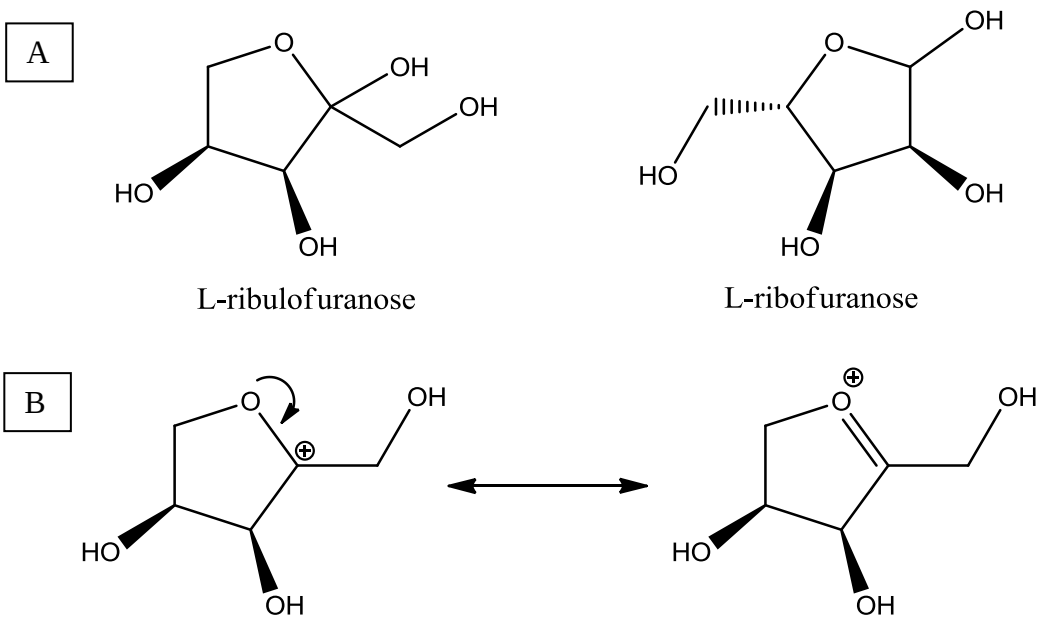


Figure 2.3. (A) The two sugars used were D-ribose and L-ribulose. (B) During condensation, the sugar is likely to go through an oxocarbenium ion transition state. Any sugar that can better stabilize oxocarbenium ion formation is expected to be more reactive towards condensation with adenine. The furanose form of ribulose would form a tertiary carbocation during a condensation reaction with adenine at the C1 position, whereas ribose would form a secondary carbocation. Since tertiary carbocations stabilized by hyperconjugation to a greater extent than secondary carbocations, ribulose is expected to be more reactive towards condensation than ribose.

A heat drying reaction between ribose and adenine under Orgel-type reaction conditions was run and chromatographed, as shown in Figure 2.4. Interestingly, several differences in the reactivity of ribulose compared to ribose were noted. First, the yield of 9-β-L-ribulofuranosyladenine was approximately the same as the yield of 9-β-D-ribofuranosyladenine when ribose and adenine were dried together in Figure 2.2. Second, the large peaks associated with ribose/adenine N6-adducts in Figure 2.2 are absent when ribose is replaced by ribulose. Finally, a peak in Figure 2.4 corresponds to the retention time and mass of 9-β-D-ribofuranosyladenine, even though no ribose was added. The peak was not NMR characterized, thus the assignment to 9-β-D-ribofuranosyladenine is still tentative. Additionally, the L-ribulose supplied by zuChem was certified to be 99.9%

pure, indicating that ribose contamination, if any, would not account for the adenosine peak. An equilibrium between ribose and ribulose (K_{eq} = 0.317) catalyzed by glucose isomerase was reported in the literature by Tewari et. al. (1985).[31] Since Orgel-type reactions reach a temperature of 100 °C, perhaps there is enough thermal energy to overcome the activation barrier of converting ribulose to ribose.

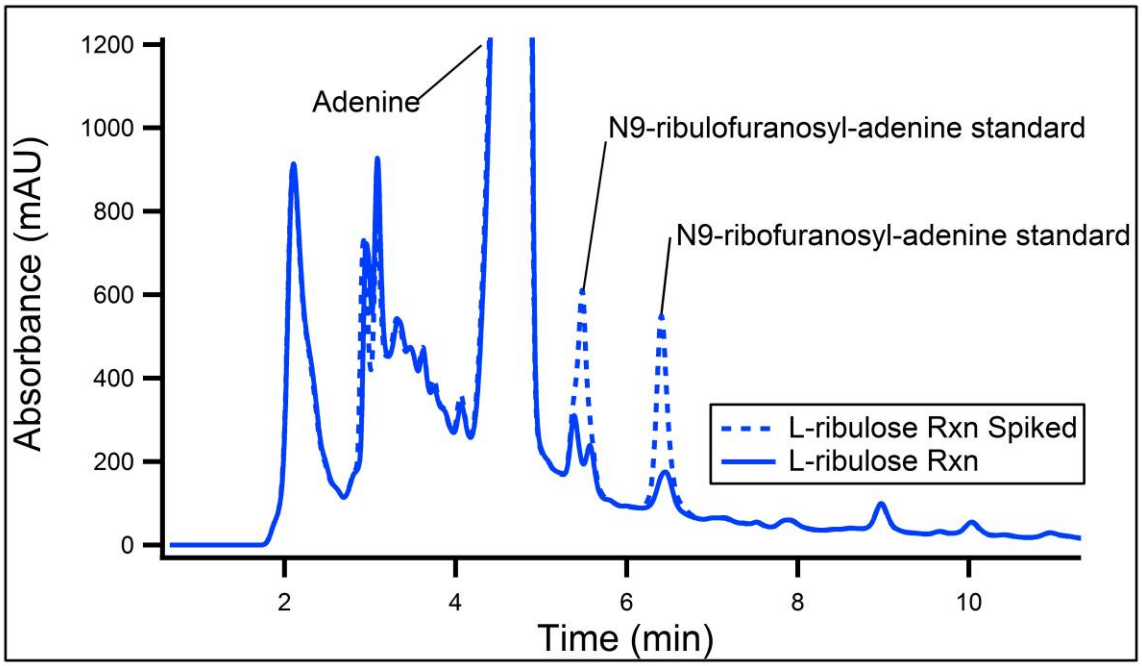


Figure 2.4. A chromatogram of an Orgel-type reaction between L-ribulose and adenine. The reaction yielded <1% 9-β-L-ribulofuranosyladenine, similar to the reaction between ribose and adenine. Additionally, adenine N6 adducts were not present in the large amounts shown in Figure 2.2. Remarkably, a peak matching the retention time and mass of 9-β-D-ribofuranosyladenine was detected. The assignment is tentative as the peak was not further characterized by NMR.

In an additional experiment shown in Figure 2.5, the N9-ribulofuranosyl adenine and the N9-ribofuranosyl adenine standard were heated under an Orgel-type reaction. When the ribulo-adenine was heated, a peak (<0.5%) corresponding with ribo-adenine was formed. Similarly, when ribo-adenine was heated, a peak (<0.5%) corresponding with ribulo-adenine was formed. The apparent ability of nucleosides to undergo sugar

isomerization in trace quantities supports the hypothesis that there exists a nontrivial equilibrium between ribulose and ribose under Orgel-type reaction conditions.

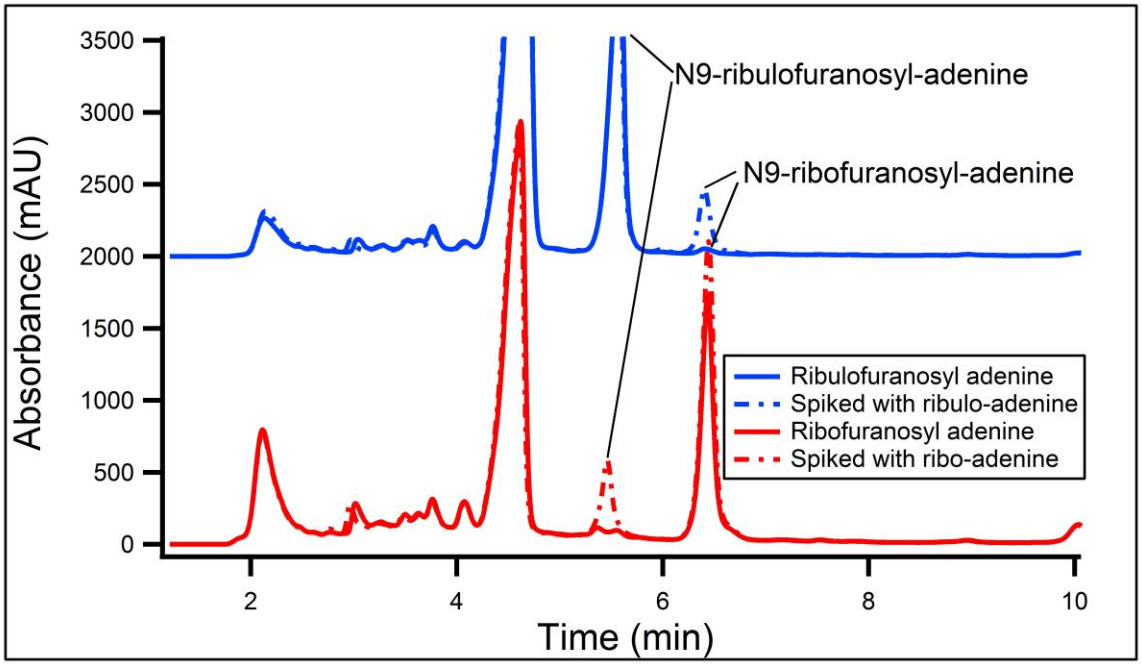


Figure 2.5. A chromatogram of an Orgel-type reaction with 5 mM 9-β-L-ribulofuranosyladenine the top trace and 5 mM 9-β-D-ribofuranosyladenine in the bottom trace. The top trace is spiked with 9-β-D-ribofuranosyladenine; the unspiked trace is overlaid with the spiked trace. Similarly, the bottom trace is spiked with 9-β-L-ribulofuranosyladenine. A trace amount of 9-β-D-ribofuranosyladenine may have been produced when heating 9-β-L-ribulofuranosyladenine.

2.4 CONCLUSIONS

L-ribulose did not add to adenine in the N9 position at greater yield compared with ribose. This result suggests that the formation of the oxonium ion is not the rate-limiting step in the condensation of a sugar and adenine. Other studies have shown that placing a good leaving group at the anomeric position increases rates of glycosylation.[32] However, significant increases in rate were only observed for reactions that occurred through an S_N1-type reaction, not for S_N2-type reactions. Since the yield of nucleosides in L-ribulose reactions seemed to be slightly lower compared to reactions with D-ribose, the glycosylation rate of adenine with L-ribulose was lower compared to D-ribose. If the

nucleosidation reaction occurs by an S_N2-type reaction, the reaction with L-ribulose would be hindered by steric bulk due to the presence of a methoxide group at the 1' position that is not present in ribose. Additionally, the large yield of potential adenine N6 adducts with ribose suggests that the N6 nitrogen is more nucleophilic compared to the N9 position in adenine. Repeated heat cycles did not increase the yield of nucleoside product for both the ribose and ribulose reactions.

However, the observation that heating ribulose with adenine creates an adenosine nucleoside with several sugar isomers, potentially including the ribofuranosyl form, is encouraging. Since ribose is not generated in high yields in the formose reaction, perhaps other sugars are able to isomerize to ribose and generate the ribofuranosyladenine nucleoside. Finally, it is interesting to note that the majority of the product between adenine and ribose consists of N6-adducts. The exocyclic amine is prone to Schiff base formation with aldehydes, such as ribose, leading to labile linkages. Since ribulose is a ketopentose, Schiff base formation is not seen between ribulose and the exocylic amine on adenine. Due to the difficulty of forming canonical N9 N,O-acetal linkages with adenine, nucleoside analogues such as Schiff base formation between exocyclic amines and aldehydes, should be investigated.

CHAPTER 3

INTERCALATOR-MEDIATED MACROMOLECULAR ASSEMBLY

3.1 INTRODUCTION

Life today depends on DNA for information storage and proteins for biosynthesis. Since RNA is capable of information storage and biosynthesis (ribozymes), origins researchers are beginning to come to a consensus that RNA-based life predated DNA-based life. Nevertheless, the chemical origins of the RNA world are still unknown. As the spontaneous synthesis of RNA in solution is highly unlikely, many groups—including the Hud lab—support the existence of a RNA-like polymer (proto-RNA) that would have acted as a transition state to the RNA world. Regardless, many obstacles exist for the abiotic synthesis of RNA or proto-RNA without the aid of protein catalysis.

The linear growth of polymers is inherently limited due to the tendency of intramolecular strand cyclization, as shown in Figure 3.1. A problem for polymer chemistry in general, strand cyclization terminates linear elongation affording small cyclized polymers. Since the backbone of RNA and DNA is relatively flexible, linear elongation of these polymers is limited to 3-4 nucleotides before the strand cyclizes. As most genes are hundreds, if not thousands of base pairs long, strand cyclization is a roadblock for the abiotic synthesis of RNA or proto-RNA. How this abiotic synthesis specifically selected for the purine and pyrimidine base pairs is another unresolved obstacle.

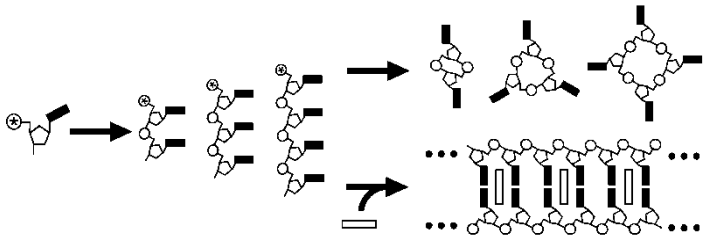


Figure 3.1. An illustration of activated oligonucleotide polymerization. The top pathway (without intercalator) leads to strand cyclization because the persistence length of single-stranded nucleic acids are 3-4 nucleotides. The bottom pathway (with intercalator) leads to linear elongation, demonstrating the effect of intercalation on the persistence length of nucleic acids.

This thesis proposes that intercalators resolve the two aforementioned obstacles (1) by increasing the persistence length of the polymer and (2) by providing a template to specifically stabilize Watson-Crick base pairs within the polymer.

3.2 EXPERIMENTAL PROCEDURES

3.2.1. Materials

All oligonucleotides were purchased from Oligos Etc. or Integrated DNA Technologies, and were subsequently purified by HPLC. Ethidium bromide was purchased from Fischer Scientific and proflavine hemisulfate from Sigma Aldrich. N-cyanoimidazole was purchased from Toronto Research Chemicals.

3.2.2. Watson-Crick paired chemical ligations

Ligation reactions were performed with d(pCGTA) (5 μM to 60 mM), $MnCl_2$ (5 mM), 10 mM triethylammonium MES (pH 6), ethidium bromide (600 μM), and N-cyanoimidazole (25 mM). Reactions were incubated at 4 °C for 72 hrs, and then ethanol

precipitated using linear polyacrylamide as a carrier for oligonucleotides as described in the literature.[33] Reactions were separated through 20% denaturing (8 M urea) PAGE (19% acrylamide:1% N, N'-methylenebisacrylamide) in TBE running buffer. Gels were stained with SYBR Gold dye purchased from Invitrogen.

3.3 RESULTS AND DISCUSSION

The system used to investigate the Watson-Crick paired intercalator-mediated assembly was a tiling 4-mer—p(CGTA)—with a two-base overhang.

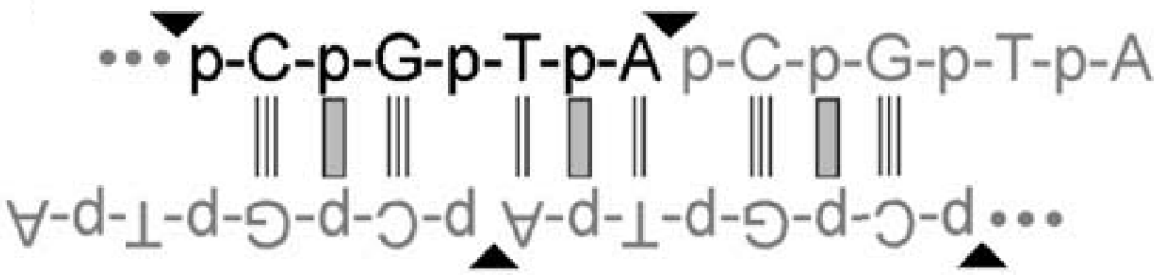


Figure 3.2. An illustration of the tiling 4-mer system used to demonstrate intercalator-mediated polymerization. d(pCGTA) is a 4-mer that has a 2-base Watson-Crick complementarity with itself, leaving a 2-base overhang.

Figure 3.3 shows that without the Watson-Crick intercalator, ethidium bromide, the only products are cyclic 4-mer and cyclic 8-mer. The reason is that the persistence length of single-stranded nucleic acids are 3-4 nucleotides. Additionally, unimolecular reactions occur at a faster rate than bimolecular reactions. When ethidium bromide was introduced to the reaction, polymerization of up to 100 nucleotides is observed, indicating that intercalation stabilizes oligonucleotide assembly and rigidifies the nucleic acid backbone.

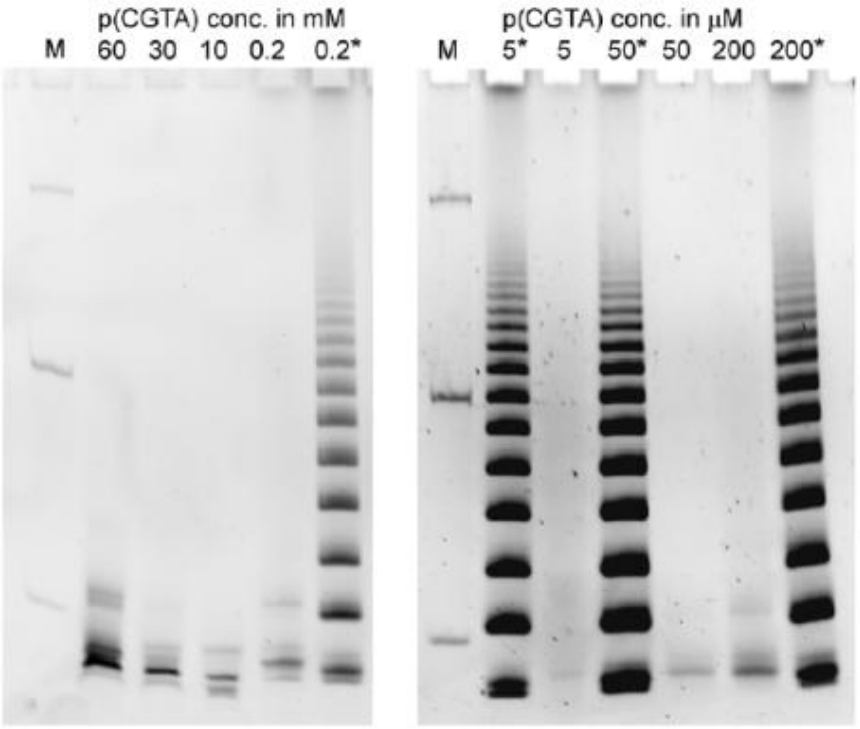


Figure 3.3. Intercalation-mediated polymerization occurs over a 12,000-fold concentration range. Numbers above gel wells indicate oligonucleotide concentration. M is a ladder consisting of 3 single-stranded oligonucleotides: 10, 32, 110 nt. The left-series of reactions demonstrate that high concentrations of oligonucleotide do not drive oligonucleotide assembly and intercalation is still necessary for polymerization even at 60 mM oligonucleotide concentration. The right-series of reactions show that intercalation-mediated polymerization is effective at low oligonucleotide concentrations. All reactions performed with 25 mM ImCN, 600 μM EtBr (*), 5 mM $MnCl_2$, and 10 mM triethylammonium MES at pH 6.0. The reactions were allowed to sit at 4 °C for ~72 hours, and then were purified by ethanol precipitation with linear polyacrylamide carrier. Gel is a 20% denaturing PAGE gel containing 8 M urea in TBE buffer.

In order to show that intercalators mediate oligomer assembly over a large concentration range, the lower and upper limits for oligonucleotide concentration were investigated. After ~72 hours of reaction at 4 ºC, the reaction mixtures were purified by ethanol precipitation. For concentrations lower than 20 μM, the purification process (ethanol precipitation) failed to obtain bands on 20% PAGE gels. Several different purification procedures were attempted, including ethanol precipitation with a linear

polyacrylamide or glycogen carrier and solid-liquid extractions. Nevertheless, these methods did not succeed in isolating the linearly elongated products.

These purifications discarded the precipitate formed during the reaction, likely insoluble byproducts of the N-cyanoimidazole reaction. By including the precipitate in the ethanol precipitation (with linear polyacrylamide as a carrier), linearly elongated products were detected to a concentration of 5 μM (tentatively detected as low as 0.5 μM), as shown in Figure 3.3. Additionally, linear elongation without intercalators did not occur for the highest oligonucleotide concentration tested (60 mM). This conclusively demonstrates the need for intercalators on at least a 12,000-fold tetranucleotide concentration range.

By demonstrating that this system can pick tiling tetranucleotides from a pool of non-tiling or interfering tetranucleotides, intercalation is able to select for Watson-Crick base pairs rather than indiscriminately assemble any oligonucleotide. Adding d(TTTT) to the reaction mixture (same conditions as above) at concentrations of 5 to 500 uM did not affect the linear elongation of the tiling tetranucleotide, as shown in Figure 3.4.

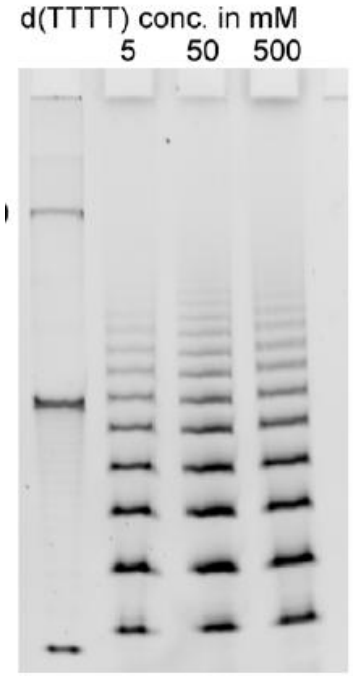


Figure 3.4. Intercalator-mediated polymerization occurs in the presence of non-tiling strands. Numbers above gel wells indicate concentration of interfering oligonucleotide d(TTTT). The figure shows that even when an interfering oligonucleotide is present at 2,500 times the amount of tiling p(CGTA), intercalator-mediated polymerization occurs. All reactions performed with 200 μM p(CGTA), 25 mM ImCN, 600 μM EtBr (*), 5 mM $MnCl_2$, and 10 mM triethylammonium MES at pH 6.0. The reactions were allowed to sit at 4 °C for ~72 hours, and then were purified by ethanol precipitation with linear polyacrylamide carrier. Gel is a 20% denaturing PAGE gel containing 8 M urea in TBE buffer.

In an effort to demonstrate that this model is compatible with multiple intercalators in the same reaction mixture, proflavine was used alongside ethidium bromide as intercalators. However, no linear elongation was observed in reaction mixtures with equimolar amounts of proflavine and ethidium bromide. As the stabilization of Watson-Crick base pairs with intercalators in solution are equilibrium reactions, this is a particularly surprising result which should be thoroughly investigated in future experiments.

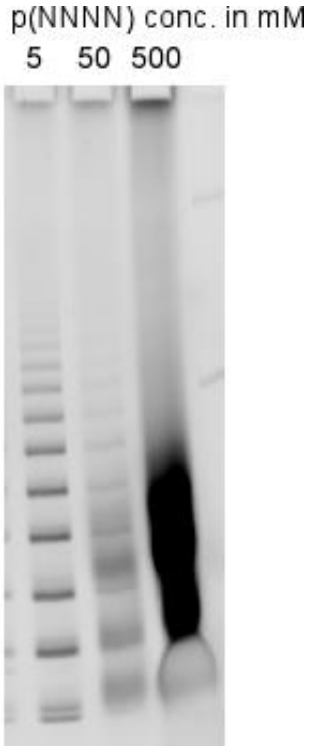


Figure 3.5. Intercalator-mediated polymerization selectively extends Watson-Crick base pairs. Numbers above gel wells indicate concentration of interfering oligonucleotide p(NNNN), where N represents a random incorporation of A, G, C or T. The difference between the experiments in Figure 3.4 and Figure 3.5 is the 5' phosphorylation of the interfering oligonucleotide. The figure shows that even when a 5' phosphorylated interfering oligonucleotide is present at 250 times the amount of tiling p(CGTA), intercalator-mediated polymerization occurs, albeit to a lesser extent. All reactions performed with 200 μM p(CGTA), 25 mM ImCN, 600 μM EtBr (*), 5 mM $MnCl_2$, and 10 mM triethylammonium MES at pH 6.0. The reactions were allowed to sit at 4 °C for ~72 hours, and then were purified by ethanol precipitation with linear polyacrylamide carrier. Gel is a 20% denaturing PAGE gel containing 8 M urea in TBE buffer.

In Figure 3.5, when p(NNNN) is added in 25-fold excess compared to the tiling 4-mer p(CGTA), intercalation showed a remarkable ability to not tolerate base-mismatches. Even at 250-fold excess, polymerization is still observed. Thus intercalation provides the necessary selectivity and energy contributions to mediate polymerization of oligonucleotides.

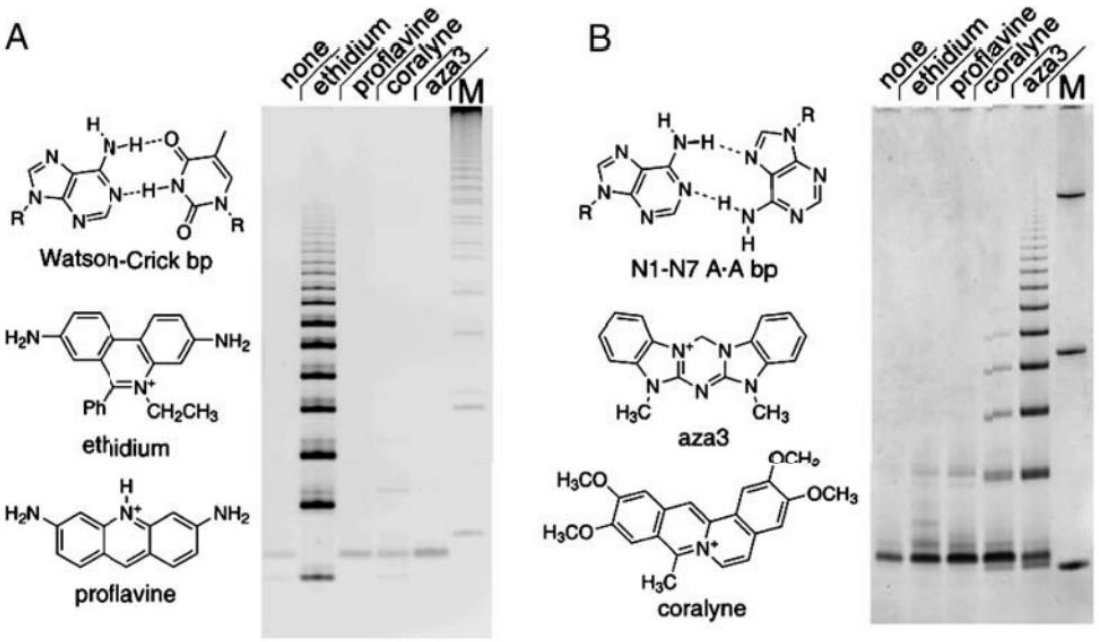


Figure 3.6. Base-pair recognition is necessary for intercalator-mediated polymerization. (A) Reactions were performed with 200 μM d(pCGTA), 25 mM ImCN, 600 μM intercalator (compound listed at top of gel), 5 mM $MnCl_2$, and 10 mM triethylammonium MES at pH 6.0. Ethidium, a Watson-Crick intercalator, mediates polymerization. However, another Watson-Crick intercalator, proflavine, does not mediate polymerization of d(pCGTA). Coralyne and aza3, homo-A duplex intercalators, do not mediate polymerization. (B) Reactions were performed with 500 μM $d(pA)_6$, 25 mM ImCN, 600 μM intercalator (compound listed at top of gel), 5 mM $MnCl_2$, and 10 mM triethylammonium MES at pH 6.0. Ethidium, a Watson-Crick intercalator, mediates polymerization. Aza3 and coralyne, to a lesser extent, mediates the polymerization of $d(pA)_6$, whereas ethidium and proflavine do not.

Finally, Figure 3.6 demonstrates that base-pair recognition by an intercalator is necessary for polymerization. Ethidium, a Watson-Crick intercalator, promotes the polymerization of d(pCGTA) and not $d(pA)_6$. The A strand forms a homo-A duplex with a base-pairing arrangement shown in Figure 3.6 (B). However, base-pair recognition is not sufficient for polymerization, as shown by proflavine's failure to polymerize d(pCGTA). This result is likely due to an incompatibility of the cyanoimidazole ligation chemistry with the binding mode of proflavine. When homo-A duplex intercalators are introduced with homo-A strands, intercalator-mediated polymerization occurs, as shown in 3.6 (B). Ethidium and proflavine do not promote the polymerization of $d(pA)_6$.

3.4 CONCLUSIONS

Since short, cyclic nucleic acids cannot store biologically relevant amounts of information and do not display catalytic ability, strand cyclization would have prevented the emergence of long polymers in the prebiotic soup. Additionally, nature selected for the Watson-Crick base pair; yet, the mechanism for that selection was unknown. Intercalation is able to mediate the non-enzymatic polymerization of oligonucleotides by circumventing strand cyclization and selecting for base-pairing types. Figure 3.3 demonstrates intercalator-mediated polymerization at a variety of concentrations. Even at 60 mM oligonucleotide concentration, where nucleic acids may self-assemble, intercalation is necessary to overcome strand cyclization. At low oligonucleotide concentrations, where assembly is less thermodynamically favorable, intercalation is able to provide sufficient binding energy to stabilize the linear form of oligonucleotides. Figures 3.4, 3.5, and 3.6 demonstrate the ability for intercalation to select for the Watson-Crick base-pair among interfering oligonucleotide strands.

This chapter addresses the problems of strand cyclization and base-pair selection but does not address how nucleic acids were non-enzymatically linked together. As such, N-cyanoimidazole, a condensing agent that is not prebiotic, was used to activate phosphorylated oligonucleotides. Finally, intercalation was unable to assembly oligonucleotides smaller than 4 nucleotides. The reason is a result of the nearest neighbor principle, where an intercalator cannot occupy the space between base-pairs if the adjacent base pair has already been intercalated. Thus, 1-, 2-, and 3-mers do not provide enough base-pairing capacity for intercalation-mediated polymerization. Future experiments will address the inability of intercalation to mediate the polymerization of

oligonucleotides less than 4 nucleotides in length, and will also address RNA intercalation.

REFERENCES

1. Crick, F., Central Dogma of Molecular Biology. *Nature* **1970,** *227* (5258), 561-563.
2. Orgel, L.; Crick, F., Anticipating an RNA world. Some past speculations on the origin of life: where are they today? *The FASEB Journal* **1993,** *7* (1), 238-239.
3. Miller, S. L., A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **1953,** *117* (3046), 528-529.
4. Oro, J., Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **1961,** *191* (4794), 1193-1194.
5. Larralde, R.; Robertson, M. P.; Miller, S. L., Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proceedings of the National Academy of Sciences* **1995,** *92* (18), 8158-8160.
6. Wachtershauser, G., Before enzymes and templates: theory of surface metabolism. *Microbiol. Mol. Biol. Rev.* **1988,** *52* (4), 452-484.
7. Morowitz, H. J.; Kostelnik, J. D.; Yang, J.; Cody, G. D., The origin of intermediary metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97* (14), 7704-7708.
8. Anet, F. A. L., The place of metabolism in the origin of life. *Current Opinion in Chemical Biology* **2004,** *8* (6), 654-659.
9. Pross, A., Causation and the Origin of Life. Metabolism or Replication First? *Origins of Life and Evolution of Biospheres* **2004,** *34* (3), 307-321.
10. Orgel, L. E., Self-organizing biochemical cycles. *Proceedings of the National Academy of Sciences* **2000,** *97* (23), 12503-12507.
11. Leslie E, O., Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in Biochemistry and Molecular Biology* **2004,** *39* (2), 99-123.
12. Santoro, S. W.; Joyce, G. F., A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences* **1997,** *94* (9), 4262-4266.
13. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P., Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **2011,** *332* (6026), 209-212.
14. Orgel, L. E., Some Consequences of the RNA World Hypothesis. *Origins of Life and Evolution of Biospheres* **2003,** *33* (2), 211-218.
15. Barks , H. L.; Buckley , R.; Grieves, G. A.; Di Mauro, E.; Hud, N. V.; Orlando, T. M., Guanine, Adenine, and Hypoxanthine Production in UV-Irradiated Formamide Solutions: Relaxation of the Requirements for Prebiotic Purine Nucleobase Formation. *ChemBioChem* **2010,** *11* (9), 1240-1243.
16. Robertson, M. P.; Miller, S. L., An efficient prebiotic synthesis of cytosine and uracil. *Nature* **1995,** *375* (6534), 772-774.
17. Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E., The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proceedings of the National Academy of Sciences* **1987,** *84* (13), 4398-4402.
18. Bean, H.; Anet, F.; Gould, I.; Hud, N., Glyoxylate as a Backbone Linkage for a Prebiotic Ancestor of RNA. *Origins of Life and Evolution of Biospheres* **2006,** *36* (1), 39-63.
19. Bean, H. D.; Sheng, Y.; Collins, J. P.; Anet, F. A. L.; Leszczynski, J.; Hud, N. V., Formation of a β-Pyrimidine Nucleoside by a Free Pyrimidine Base and Ribose in a

Plausible Prebiotic Reaction. *Journal of the American Chemical Society* **2007,** *129* (31), 9556-9557.
20. Ingar, A.-A.; Luke, R. W. A.; Hayter, B. R.; Sutherland, J. D., Synthesis of Cytidine Ribonucleotides by Stepwise Assembly of the Heterocycle on a Sugar Phosphate. *ChemBioChem* **2003,** *4* (6), 504-507.
21. Sanchez, R. A.; Orgel, L. E., Studies in prebiotic synthesis , : V. Synthesis and photoanomerization of pyrimidine nucleosides. *Journal of Molecular Biology* **1970,** *47* (3), 531-543.
22. Fuller, W. D.; Sanchez, R. A.; Orgel, L. E., Studies in prebiotic synthesis , : VI. Synthesis of purine nucleosides. *Journal of Molecular Biology* **1972,** *67* (1), 25-33.
23. Fuller, W. D.; Sanchez, R. A.; Orgel, L. E., Studies in prebiotic synthesis. VII. *Journal of Molecular Evolution* **1972,** *1* (3), 249-257.
24. Maurel, M.-C.; Convert, O., Chemical structure of a prebiotic analog of adenosine. *Origins of Life and Evolution of Biospheres* **1990,** *20* (1), 43-48.
25. Sheng, Y.; Bean, H. D.; Mamajanov, I.; Hud, N. V.; Leszczynski, J., Comprehensive Investigation of the Energetics of Pyrimidine Nucleoside Formation in a Model Prebiotic Reaction. *Journal of the American Chemical Society* **2009,** *131* (44), 16088-16095.
26. Powner, M. W.; Gerland, B.; Sutherland, J. D., Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **2009,** *459* (7244), 239-242.
27. Šponer, J. E.; Šponer, J.; Fuentes-Cabrera, M., Prebiotic Routes to Nucleosides: A Quantum Chemical Insight into the Energetics of the Multistep Reaction Pathways. *Chemistry – A European Journal* **2011,** *17* (3), 847-854.
28. Robertson, M. P.; Joyce, G. F., The Origins of the RNA World. *Cold Spring Harbor Perspectives in Biology* **2010**.
29. Decker, P.; Schweer, H.; Pohlamnn, R., Bioids : X. Identification of formose sugars, presumable prebiotic metabolites, using capillary gas chromatography/gas chromatography--mas spectrometry of n-butoxime trifluoroacetates on OV-225. *Journal of Chromatography A* **1982,** *244* (2), 281-291.
30. Zubay, G.; Mui, T., Prebiotic Synthesis of Nucleotides. *Origins of Life and Evolution of Biospheres* **2001,** *31* (1), 87-102.
31. Tewari, Y. B.; Goldberg, R. N., An investigation of the equilibria between aqueous ribose, ribulose, and arabinose. *Biophysical Chemistry* **1985,** *22* (3), 197-204.
32. Denekamp, C.; Sandlers, Y., Formation and stability of oxocarbenium ions from glycosides. *Journal of Mass Spectrometry* **2005,** *40* (8), 1055-1063.
33. Gaillard, C.; Strauss, F., Ethanol precipitation of DNA with linear polyacrylamide as carrier. *Nucleic Acids Research* **1990,** *18* (2), 378.